Dear author,

Please note that changes made in the online proofing system will be added to the article before publication but are not reflected in this PDF.

We also ask that this file not be used for submitting corrections.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

CHAPTER

5

c0005

Computational models of multisensory integration

David Meijer, Uta Noppeney

Computational Cognitive Neuroimaging Laboratory, Computational Neuroscience and Cognitive Robotics Centre, University of Birmingham, Birmingham, United Kingdom

s0010

Introduction

p0015 Various sensory organs continuously provide our brains with uncertain information about our environment. Critically, every sensor has its specific limitations. For example, the sensitivity of our eyes' photoreceptors is optimized for use during daylight (e.g., photoreceptor sensitivity of nocturnal insects is much higher[1]). Our ears are specialized for detecting differences in sound pitch, but they provide only imprecise estimates for the location of a sound's source.

p0020 Imagine you are in a dimly lit bedroom at night and you hear the sound of a mosquito. To obtain the most precise estimate of the mosquito's location, the brain should combine uncertain spatial information furnished by the auditory and visual senses. Critically, the brain should integrate sensory signals only when they pertain to the same event, but process them independently when they come from different events. For example, we are all familiar with those vague black spots on the wall that look annoyingly like mosquitos in the dark. These immobile black spots should not be integrated with the mosquito's buzzing sound around the head. In short, to generate a coherent percept of the environment, the brain needs to infer whether or not sensory signals are caused by common or independent sources. This process has been termed multisensory causal inference.[2]

p0025 In this chapter, we will explore the computational operations that the brain may use to solve these two challenges involved in multisensory perception, i.e., (1) how to weight and integrate signals that come from a common source into a unified percept and (2) how to infer whether signals come from common or independent sources.

p0030 In the first section, we will introduce the normative Bayesian framework focusing on perception based on input from a single sensory channel and prior expectations. In the second section, we will describe how the brain integrates signals from multiple sensory channels pertaining to the same event into a unified percept (i.e., so-called forced fusion model). In the

https://doi.org/10.1016/B978-0-12-812492-5.00005-X

Copyright © 2020 Elsevier Inc. All rights reserved.

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

third section, we will explore the more general case of multisensory perception in the face of uncertainty about the world's causal structure, i.e., uncertainty about whether signals are caused by common or independent sources. Hence, this final case combines the two challenges facing the brain in a multisensory world: causal inference and weighted sensory integration. Each section first describes the normative Bayesian model and then briefly reviews the empirical evidence that shows the extent to which data from human or nonhuman primates are in accordance with those computational principles.

s0015 Combining information from a single sensory channel with prior knowledge

p0035 Any sensory signal that reaches the cerebral cortex is inevitably contaminated with various sources of noise. Let us consider how an observer can estimate the location of an event for spatial orienting from visual inputs. An observer's eyes are bombarded with photons, and each eye's lens refracts the photons such that a ray of focused light hits the retina. There, photoreceptors and ganglion cells transform the electromagnetic radiation into action potentials. This eventually, via several synapses, results in an activity pattern in the visual cortex. Importantly, noise may be introduced at each of those processing stages. The eye's view can be partially obscured by a dirty window, and its lens is unlikely to be perfectly in focus; the transformation from photons to action potentials functions in bulk[3]; and synaptic transmission is a probabilistic process.[4] In short, the sensory organs and systems provide the brain only with an uncertain or noisy estimate of a particular property (e.g., spatial location) of events and objects in the outside world.

p0040 To constrain perceptual inference, the observer can combine the noisy sensory evidence with *prior* knowledge or expectations. For example, in our natural environment, it is very unlikely to observe a concave human face, where the tip of the nose faces away from the observer. When an observer is shown the inside of a mask, the brain often falsely interprets the image such that the nose is perceived to be facing the observer. The visual hollow-face illusion, as this effect was dubbed, is only one of many examples where prior knowledge affects our perception.[5]

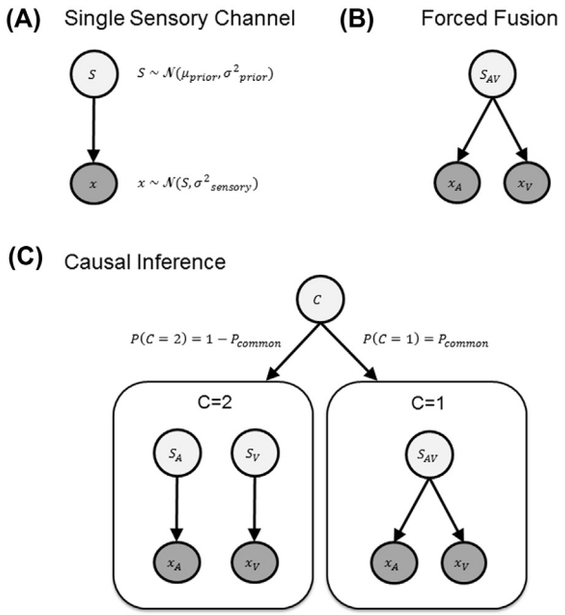
p0045 The normative Bayesian framework in neuroscience posits that the brain forms a probabilistic generative model of the sensory inputs that is inverted during perceptual inference (= recognition model). Bayesian probability theory offers a precise formulation of how observers should combine uncertain information such as different sorts of noisy sensory evidence and prior knowledge to form the most reliable representation of the world. It thus sets a benchmark of a so-called "ideal observer" or optimal performance given a particular loss function against which an organism's neural and behavioral responses can be compared.

p0050 Fig. 5.1A shows the graphical model that illustrates the generative process for the spatial localization example above based on a single sensory channel and prior knowledge. A hidden source at the true location S generates a noisy sensory signal representation X. The true location S is sampled from a prior distribution, which is often assumed to be a Gaussian with mean μ: $S \sim N(\mu_{prior}, \sigma^2_{prior})$. The sensory signal is corrupted by noise, i.e., sampled from a Gaussian centered on the true source location: $x \sim N(S, \sigma^2_{sensory})$. The generative model defines the probability of each sensory input given a particular source location $P(x|S)$. During perception, the observer needs to invert this generative model to compute the posterior probability $P(S|x)$, i.e., the probability of a spatial location given the sensory input x, by combining sensory evidence and prior knowledge. According to Bayes' rule, the posterior probability of

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.



f0010 FIGURE 5.1 **Generative models corresponding to the three different cases.** (A) Single sensory signal: a hidden source generates a sensory signal that is corrupted by noise. (B) Forced fusion: a hidden source generates two sensory signals (e.g., auditory and visual) that are independently corrupted by noise. (C) Causal inference model explicitly models the potential causal structures that could have generated the two sensory signals (e.g., auditory and visual). In the full segregation model component (left), two independent hidden sources generate the auditory and visual signals. In the forced fusion model component, a common source generates two sensory signals (e.g., auditory and visual). A Bayesian causal inference estimate combines the estimates obtained from those two model components using a specific decision function (e.g., model averaging). *Adapted from Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception.* PLoS One. *2007;2(9):e943.*

a spatial location given a particular sensory input, $P(S|x)$, is proportional to the product of the likelihood $P(x|S)$ and the prior $P(S)$:

$$P(S|x) = \frac{P(x|S) * P(S)}{P(x)} \propto P(x|S) * P(S) \tag{5.1}$$

p0055 The normalization constant $P(x)$ can be obtained from the product of the likelihood function and the prior by marginalizing (i.e., integrating) over all possible locations S:

$$P(x) = \int P(x|S) * P(S) * dS \tag{5.2}$$

p0060 The observer then needs to minimize a particular loss function that specifies the cost of selecting the estimate $\widehat{S}$ given the true location S to report a final point estimate. For instance, using the squared error loss function, the observer would report the mean of the posterior distribution as the final spatial estimate. By contrast, using a zero-one loss function, the observer

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

would report the maximum a posteriori estimate (MAP), i.e., the mode of the posterior distribution. Critically, under Gaussian assumptions of both prior and likelihood, the posterior mean and mode are identical, i.e., both loss functions yield the same final estimate. However, asymmetric posterior distributions lead to different estimates for the posterior mean and MAP.[6,7]

p0065 Priors can emerge at multiple timescales potentially ranging from seconds to evolutionary times. For instance, during evolution, certain hardwired neural priors may have emerged as a result of selection pressures.[8] Likewise, other hardwired priors may be fine-tuned during neurodevelopment when the immature brain is exposed to the statistics of the sensory inputs.[9] Finally, the brain is thought to rapidly adjust priors to changes in the input statistics across and perhaps even within trials where the posterior of the current trial or time point forms the prior for the next trial or time point.[10,11] Priors are critical to constrain perceptual inference in the face of uncertainty resulting from noise, occlusion, etc. As we will derive in greater detail in the next "forced fusion" section, the influence of the prior on the final posterior estimate should be greatest if the sensory input is noisy and uncertain. This is because different sorts of evidence (e.g., prior vs. sensory evidence or different sensory evidences) should be combined in a manner weighted by their relative reliabilities (see Forced fusion: integrating sensory signals that come from a common source section for details).

p0070 Priors can be formed about all sorts of properties such as spatial location, shape, speed, etc. Indeed, numerous studies have demonstrated how prior knowledge or expectations shape and *bias* perceptual inference in our natural environment or designed experimental settings: the light-from-above prior (objects with ambiguous depth seem to face forward if the shadow is below them[12]), the circularity assumption (we tend to think that an object's depth is equal to its width[13]), the foveal bias (relevant objects are more likely to appear in the center of our field of view[14,15]), the slow speed preference (most objects do not move or tend to move slowly[16,17]), and the cardinal orientation prior (vertical and horizontal orientations can be more frequently found[18]). In the latter example, the experimentally determined probabilities of the human prior distribution for orientations were shown to match the environmental statistics for orientations that were found in a large set of photographs.[18] In addition to the long-term priors, the brain can also rapidly adapt priors to the dynamics of statistical regularities. In laboratory experiments, participants may learn the distribution from which the stimuli are sampled (e.g., the range of stimulus durations in a time-interval estimation task[19]). In the real world, they can adopt prior distributions that apply to a particular situation (e.g., the typical velocities for a ball in a game of tennis[20]). Multiple studies have also shown that the biasing influence of the prior is—as expected (see above)—inversely related to the reliability of the sensory stimuli.[16–20]

p0075 At the neural level, a recent functional magnetic resonance imaging (fMRI) study has shown that the brain estimates the reliability or precision of sensory representations in primary visual cortex (V1) on a trial-by-trial basis.[21] Participants were presented with visual gratings that varied in their orientation across trials. On each trial, they indicated the perceived orientation using a rotating bar. Critically, even though no external noise was added to the stimuli, the precision of sensory representations in V1 may vary across trials because of internal neural noise. Indeed, the uncertainty estimated from the activity patterns in the visual cortex varied across trials. Moreover, it correlated positively with the variance of participants' responses and negatively with their orientation errors. The results of this study[21] suggest that sensory cortices represent stimulus uncertainty on a trial-by-trial basis and that this uncertainty affects behavioral performance, as predicted by probabilistic models of Bayesian inference.

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

s0020 Forced fusion: integrating sensory signals that come from a common source

p0080 Many events and objects in the natural environment can be perceived concurrently by multiple senses that are each specialized for specific features of the outside world. Signals from different senses can provide complementary information. For instance, honey can be perceived as yellow by vision, but tastes sweet. Alternatively, multiple senses can provide redundant information about the same physical property such as spatial location. Thus, we can locate a puncture in a bicycle's inner tube by vision, audition, or touch (i.e., seeing, hearing, or feeling where the air flows out of the tube). In the case of redundant information across the senses, multisensory perception enables the observer to form a more precise or reliable (reliability being the inverse of variance) estimate of the environmental property in question by integrating evidence across the senses.

p0085 Fig. 5.1B shows the generative model for spatial localization based on redundant auditory and visual information. The generative model assumes one single source at the true location S_{AV} that emits two internal sensory signals; in this case, a visual and an auditory signal: x_A and x_V. As we do not allow for the two signals to be generated by two independent sources, we refer to this generative model as the forced fusion scenario, where optimal performance can be obtained by mandatory sensory integration. Again, as in the unisensory case, we assume that the auditory and visual signals, x_A and x_V, are corrupted by independent Gaussian noise. Hence, we sample x_A and x_V independently according to $x_A \sim \mathrm{N}(S_{AV}, \sigma^2{}_A)$ and $x_V \sim \mathrm{N}(S_{AV}, \sigma^2{}_V)$.

p0090 During perceptual inference, the observer needs to compute the posterior probability of the spatial location given auditory and visual inputs according to Bayes' theorem:

$$P(S_{AV}|x_A, x_V) = \frac{P(x_A, x_V|S_{AV}) * P(S_{AV})}{P(x_A, x_V)} \propto P(x_A, x_V|S_{AV}) * P(S_{AV}) \tag{5.3}$$

p0095 Furthermore, as auditory and visual inputs are assumed to be conditionally independent (i.e., independent noise assumption across sensory channels), we can factorize the likelihood:[22]

$$P(S_{AV}|x_A, x_V) \propto P(x_A|S_{AV}) * P(x_V|S_{AV}) * P(S_{AV}) \tag{5.4}$$

p0100 Furthermore, most studies in multisensory integration assume an uninformative or flat prior $P(S_{AV})$, where we can ignore the influence of the prior. As a result, the maximum a posteriori estimate turns into a maximum likelihood estimate:

$$P(S_{AV}|x_A, x_V) \propto P(x_A|S_{AV}) * P(x_V|S_{AV}) \tag{5.5}$$

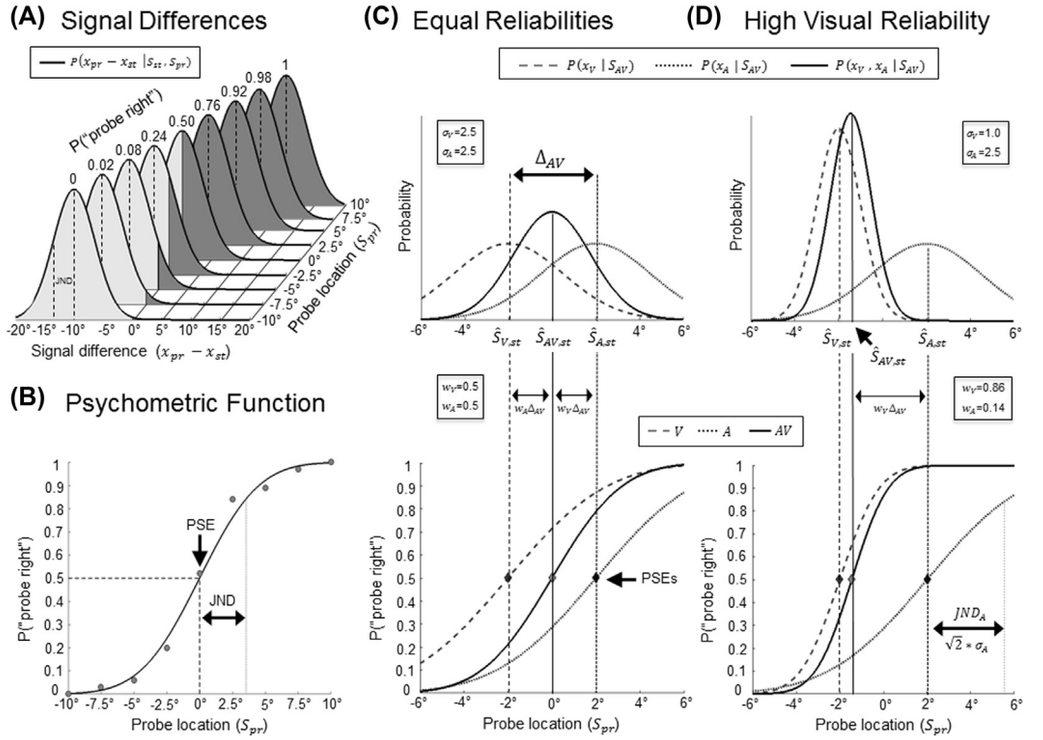
p0105 Assuming independent Gaussian noise and uninformative priors, the optimal, most precise (i.e., most reliable or with minimum variance) audiovisual estimate $\widehat{S}_{AV}$ can be computed as a reliability-weighted linear average of the two unisensory estimates[22,23]:

$$\widehat{S}_{AV} = w_A \widehat{S}_A + w_V \widehat{S}_V \quad \text{with } w_A = \frac{r_A}{r_A + r_V} \quad \text{and} \quad w_V = \frac{r_V}{r_A + r_V} = 1 - w_A \tag{5.6}$$

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.



f0015 **FIGURE 5.2** **Forced fusion model, maximum likelihood estimation, and psychometric perturbation analysis.** (A) Signal detection theoretic analysis of a 2IFC spatial discrimination task. For each true probe stimulus location S_{pr} (and standard stimulus location S_{st} at 0 degree), the observer computes a spatial estimate of the probe signal (x_{pr}) relative to the standard signal (x_{st}): i.e., the spatial signal difference $x_{pr} - x_{st}$. Because of trial-specific external and internal noise affecting both standard and probe stimuli, the signal difference is assumed to vary from trial to trial for identical true stimuli locations, S_{pr} and S_{st}, according to a Gaussian probability distribution with a standard deviation of $\sqrt{2} * \sigma_{sensory}$ that defines the summed sensory noise of the standard and probe stimuli. The observer provides a "probe right" discrimination response when the spatial signal difference is greater than zero degrees visual angle (i.e., $x_{pr} - x_{st} > 0°$). (B) Psychometric function. For the data of panel A, a cumulative Gaussian shows the probability (or fraction of trials; *gray circles*, including measurement noise) of "probe right" responses as a function of the true probe location S_{pr}. The probability "probe right" (in B) corresponds directly to the integral (i.e., dark shaded area in (A) of the Gaussian probability distribution (in A) where $x_{pr} - x_{st} > 0°$). The point of subjective equality (PSE) refers to the probe location associated with $P(\text{"}probe\ right\text{"}) = 0.5$. The just noticeable difference (JND) refers to the difference in probe stimulus locations at the two thresholds: $P(\text{"}probe\ right\text{"}) = 0.5$ and $P(\text{"}probe\ right\text{"}) \approx 0.84$. In a 2IFC task, the JND (in B) is equal to the standard deviation of the Gaussian probability distribution of signal differences (in A): *i.e.* $JND = \sqrt{2} * \sigma_{sensory}$. C and D. Maximum likelihood estimation (MLE) under forced fusion assumptions: the observer is presented with an audiovisual conflict stimulus (Δ_{AV}), i.e., the visual signal is presented at $-\frac{1}{2}\Delta_{AV}$ and the auditory signal is presented at $+\frac{1}{2}\Delta_{AV}$, as the standard in the first interval and an audiovisual congruent stimulus as the probe in the second interval. The Gaussians (top) show the likelihood functions and unbiased spatial estimates (i.e., maximum likelihood estimates; *vertical lines*) from the standard stimulus separately for the visual signal ($x_V = S_{V,st} = -\frac{1}{2}\Delta_{AV}$, *dashed*), the auditory signal ($x_A = S_{A,st} = +\frac{1}{2}\Delta_{AV}$, *dotted*), and the combined audiovisual signal as obtained from MLE-based integration (Eqs. 5.6 and 5.7, *solid*). The means of the Gaussian likelihood functions for the audiovisual conflict stimuli (top) can be estimated as the PSEs of the cumulative Gaussians (bottom) obtained from auditory, visual, and audiovisual 2IFC trials where the audiovisual spatial conflict stimulus is presented as the standard stimulus (i.e., see above $S_{st} = \pm\frac{1}{2}\Delta_{AV}$) and the probe stimulus is presented at variable degrees

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

where the reliability is defined as the inverse of the Gaussian's variance: $r = \frac{1}{\sigma^2}$. Moreover, the reliability of this audiovisual estimate can be expressed as the sum of the two unisensory reliabilities:

$$r_{AV} = r_A + r_V \text{ which is equivalent } \sigma^2_{AV} = \frac{\sigma^2_A * \sigma^2_V}{\sigma^2_A + \sigma^2_V} \tag{5.7}$$

p0110 Hence, the reliability of the audiovisual estimate is greater than (or equal to) the maximal reliabilities of the unisensory estimates. Eq. (5.7) shows formally that multisensory integration increases the precision of the percept. The maximal multisensory variance reduction by a factor of 2 can be obtained when the variances of the two sensory signals are equal.

p0115 In summary, the maximum likelihood estimation (MLE) model under forced-fusion assumptions makes two critical predictions for human multisensory perception performance. First, the variance associated with the multisensory percept is smaller than (or equal to) the minimal variance of the unisensory percepts (Eq. 5.7). Second, the multisensory percept is obtained by integrating sensory inputs weighted by their relative reliabilities (Eq. 5.6).

p0120 In the following, we will describe the standard psychophysical approach[23,24] that allows us to test whether human behavior is in accordance with these two MLE predictions. The main steps for testing each of the two MLE predictions involve (1) estimating the unisensory variances from perceptual performance on unisensory trials, (2) using Eqs. (5.6) and (5.7) to make parameter-free MLE predictions about the multisensory variance and the sensory weights applied during multisensory integration, and (3) comparing these predictions with the multisensory variances and weights empirically measured during multisensory perceptual performance. We will use an audiovisual spatial discrimination task as an example.[25]

p0125 To investigate whether audiovisual integration of spatial inputs leads to the MLE-predicted variance reduction, we need to measure the variances associated with auditory, visual, and audiovisual percepts. The empirical variances for these percepts (e.g., spatial estimates) can be estimated from participants' responses in a two-interval forced choice (2IFC) paradigm. On each trial, the observer is presented with a standard stimulus in the first interval at zero degrees ($S_{st} = 0°$) and a probe stimulus in the second interval at variable degrees of visual angle along the azimuth (S_{pr}). Standard and probe stimuli are both presented in the visual, auditory, or audiovisual modalities. The observer discriminates whether the probe stimulus is on the left or right side of the standard. Next, we fit psychometric functions, i.e., a cumulative Gaussian (ψ), to the percentage "perceived right" responses as a function of the visual angle of the presented probe separately for the visual, auditory, and audiovisual conditions (e.g., using MLE for fitting[26]; see Fig. 5.2A and B).

of visual angle. (C) For equal visual and auditory reliabilities, the means of the Gaussian likelihood functions and the PSEs of the corresponding cumulative Gaussian psychometric functions are equal to the average of the auditory and visual means or PSEs. (D) If the visual reliability is greater (i.e., visual variance is smaller) than the auditory one, the visual signal is assigned a greater weight. As a result, the mean of the audiovisual estimate is closer to the visual than the auditory estimate. As shown in the figure, we can estimate the sensory weights from the PSEs of the psychometric functions of the unisensory visual, unisensory auditory, and audiovisual conflict conditions in a 2IFC task. *Adapted from Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion.* Nature. *2002;415(6870):429–433.*

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

$$\psi(S_{pr}) = \frac{\beta}{\sqrt{2\pi}} \int_{-\infty}^{S_{pr}} \exp\left(-\frac{\beta^2 (S_{pr} - \alpha)^2}{2} \right) \tag{5.8}$$

where α is the point of subjective equality (PSE), i.e., the probe location where the psychometric function equals 0.5, and it is equally likely for the observer to perceive the probe left or right of the standard. Furthermore, the just noticeable difference (JND), i.e., the difference in probe locations between the PSE and the point where the psychometric function equals ~0.84, is given by $\frac{1}{\beta}$. Importantly, as shown in Fig. 5.2A and B, the PSE and JND obtained from the psychometric function as a cumulative Gaussian correspond directly to the mean (μ) and standard deviation (σ) of the Gaussian distribution that describes the perceptual noise for the auditory, visual, or audiovisual spatial estimates.[27] More specifically, as we used a 2IFC paradigm in which sensory noise of both standard and probe contribute equally to the signal differences ($x_{pr} - x_{st}$), we can compute the perceptual variance for the auditory, visual, and audiovisual conditions from the JNDs of their psychometric functions according to $JND^2 = 2\sigma^2$. Using Eq. (5.7), we can then assess whether the empirically measured AV variance is in accordance with the MLE-predicted AV variance computed from the unisensory auditory and visual variances.

p0130 To investigate whether observers integrate sensory signals weighted by their relative reliabilities as predicted by MLE, we use a so-called perturbation analysis.[28] For the perturbation analysis, we need to introduce a small nonnoticeable conflict between the auditory and visual signals of the audiovisual standard stimulus (n.b. no audiovisual conflict is introduced for the probe stimulus). For instance, we can shift the auditory signal by $+\frac{1}{2}\Delta_{AV}$ and the visual signal by $-\frac{1}{2}\Delta_{AV}$ relative to $S_{AV,st}$ congruent (=0°). If the auditory and visual signals are equally reliable and hence equally weighted in the AV spatial estimate, the perceived AV location of the conflict AV stimulus is equal to the perceived location of the corresponding congruent AV stimulus (see Fig. 5.2C, top panel). Yet, if the visual reliability is greater than the auditory reliability, the perceived location (i.e., spatial estimate) for the AV conflict stimulus should be biased toward the true location of the visual signal (i.e., in the above case shifted toward the left; see Fig. 5.2D, top panel) and vice versa for greater auditory reliability. The more frequently reported visual bias on the perceived sound location has been coined the ventriloquist effect, a perceptual illusion known since ancient times. Yet, the opposite bias from audition to vision can also emerge if the visual signal is rendered less reliable.[25] To summarize, the crossmodal bias operating from vision to audition and vice versa provides us with information about the relative sensory weights applied during multisensory integration. Formally, we can quantify the weights applied to the auditory and visual signals from the PSEs of the psychometric functions obtained from the AV conflict conditions by rewriting Eq. (5.6) (see Fig. 5.2C and D, lower panels):[23]

$$w_{A,emp} = \frac{PSE_{\Delta AV} - S_{V,st}}{S_{A,st} - S_{V,st}} \quad \text{with } w_V = 1 - w_A \tag{5.9}$$

p0135 Note that this equation implicitly assumes that unisensory auditory and visual perception are unbiased (i.e., the PSEs of the unisensory psychometric functions are equal to zero). These

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

empirical sensory weights can then be statistically compared with the MLE-predicted weights computed from the JNDs of the unisensory psychometric functions according to Eq. (5.6).

p0140 Critically, measuring the sensory weight requires a difference in the location of unisensory component signals, i.e., the presentation of incongruent audiovisual signals. While a greater intersensory conflict may enable a more reliable estimation of sensory weights, it progressively violates the forced fusion assumption and makes it less likely that observers assume a common source for the sensory signals. As a rule of thumb, a Δ_{AV} equal to the JND of the more reliable sensory signal has been proposed to be adequate.[24]

p0145 Numerous psychophysical studies have suggested that human observers integrate two sensory signals near-optimally, i.e., as predicted by the forced fusion model outlined above. For instance, near-optimal integration has been shown for visual-tactile size estimates in a seminal study by Ernst and Banks.[23] Four participants judged, by looking and/or feeling, whether the height of a raised ridge stimulus was taller than a standard comparison height. The true height of the ridge varied with small deviations from the standard height on a trial-by-trial basis. The used apparatus allowed the researchers to independently decrease the visual reliability by addition of visual noise at four different levels. Psychometric functions were fit to the unisensory and bisensory responses such that MLE-predicted and empirical weights and variances could be compared (as described above). Results indicated that the visual variance increased and visual weights decreased with increasing visual noise levels (as predicted by Eq. 5.6). Importantly, the empirical visual weights and visual-haptic variances were similar to the MLE-predicted weights and variances for all four noise levels (with a notably clear bisensory variance reduction when the visual and haptic perceptual reliability were similar; Eq. 5.7); thus suggesting that visual and haptic sensory signals were integrated in (near-) optimal fashion.[23] A follow-up experiment by the same group, using similar stimuli and apparatus, replicated the finding of an optimal variance reduction for visual-tactile size estimates (in conditions with negligible spatial disparity between the two sensory-specific cues).[29] Other examples of multisensory integration for which human behavior was shown to be in line with MLE include audiovisual location estimates,[25] audiovisual frequency discrimination,[30,31] visual-tactile object-shape judgments,[32] audiovisual duration estimates,[33] and audiovisual motion-speed discrimination.[34]

p0150 At the neural level, neurophysiological studies in nonhuman primates have shown that neural populations[35] and single neurons[36,37] integrate sensory signals weighted by their reliabilities in line with MLE predictions in visual-vestibular motion discrimination tasks. Furthermore, Fetsch et al.[35] showed that the variances and sensory weights obtained from decoding spiking rates in a population of multisensory neurons were qualitatively comparable with the variances and weights observed at the behavioral level. At a more implementational level, these authors have proposed the divisive normalization model.[38,39] This normalization model mediates reliability-weighted sensory integration, because the activity of each neuron is normalized by the activity of the entire pool of neurons.

p0155 Additional evidence in support of reliability-weighted multisensory integration at the neural level comes from several human fMRI studies showing that the connectivity between unisensory regions and association regions such as the superior temporal sulcus depends on the relative audiovisual reliabilities in speech recognition tasks.[40,41] Likewise, the blood oxygenation level-dependent response induced by somatosensory inputs in parietal areas was modulated by the reliability of concurrent visual input during a visuohaptic size discrimination task.[42]

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

p0160 Despite considerable evidence in support of MLE-optimal integration in human and nonhuman primates, accumulating research has also revealed situations where the sensory weights and reduction in multisensory variance are not fully consistent with the predictions of MLE. These findings highlight assumptions and limitations of the standard MLE forced fusion model for multisensory perception.

p0165 Focusing on the sensory weights, numerous studies have shown that human observers overweight a particular sensory modality in a range of tasks. Most prominently, in contrast to the classical study by Alais and Burr[25] showing MLE-optimal auditory and visual weights in spatial localization, Battaglia et al.[43] reported that observers rely more strongly on visual than auditory signals for spatial localization. Likewise, a series of studies have shown auditory overweighting in audiovisual temporal judgment tasks,[44,45] vestibular overweighting in visual-vestibular self-motion tasks,[46,47] visual overweighting in a visual-vestibular self-rotation task,[48] and haptic overweighting in a visual-haptic slant discrimination task.[49] In all of those studies, the sensory modality that is overweighted was the modality that is usually more reliable for this particular task in everyday experiences. One may therefore argue that the brain adjusts the weights of the sensory inputs not only based on the input's current reliability but also imposes a modality-specific reliability prior that reflects the modality's reliability for a particular property or task in everyday life.[43,45]

p0170 With respect to the second MLE prediction of multisensory variance reduction, numerous studies, covering a variety of sensory modalities and tasks, have also shown a decrease in multisensory variance that is smaller than predicted by the forced fusion model (Eq. 5.7). For example, this was shown for audiovisual interval duration judgments,[44] audiovisual speed discrimination,[50] visual-haptic slant discrimination,[49] and visual-haptic size and depth estimation tasks.[51,52] This "suboptimal" integration performance can be explained by several key assumptions of the forced fusion model that may not hold in our natural environment. First, the forced fusion model assumes that two signals are necessarily generated by one single source. However, in the real world, sensory signals can be generated by common or independent sources, leading to uncertainty about the world's causal structure (see next section). Likewise, in some experimental settings, the observer may take into account this causal uncertainty, in particular if conflict trials are included or artificial stimuli are used that do not enhance the observer's forced fusion or common source assumptions.[50,51] Second, the MLE model assumes that the sensory noise is independent between sensory modalities.[22] This assumption may be violated in some multisensory estimation tasks where dependencies exist between sensory modalities as a result of crossmodal adaptive calibration (e.g., auditory spatial estimates can be recalibrated by synchronous visual signals through a process that is different from multisensory integration).[51,53–56] Third, the MLE model does not include additional sources of noise that may be added after integration, e.g., during decision-making and response selection.[44,52]

s0025 Causal inference: accounting for observer's uncertainty about the world's causal structure

p0175 The forced fusion model presented in the previous section accommodates only the special case of where two signals come from a common source. As a result, it can only model that

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

two signals are integrated in a mandatory fashion. Yet, in our natural environment, our senses are bombarded with many different signals. In this more naturalistic scenario, an observer should bind signals into one coherent and unified percept only when they come from a common source, but he needs to treat them separately when they come from independent sources. Critically, the observer does not know the causal structure underlying the sensory signals. Instead, he needs to infer whether signals come from common or independent sources from the signals themselves. A range of correspondence cues such as temporal coincidence and correlations, spatial colocation, and higher-order cues such as semantic, phonological, metaphoric, etc., correspondences (see Chapter 11)[57–69] are critical cues informing observers about whether signals come from a common source and should thus be integrated. Hence, multisensory perception in our natural environment relies on solving the so-called causal inference problem.[2] It requires observers not only to deal with uncertainty about perceptual estimates but also with causal uncertainty, i.e., their uncertainty about the world's causal structure.

p0180 Spatial ventriloquism is a prominent audiovisual perceptual illusion that illustrates not only reliability-weighted integration (see Forced fusion: integrating sensory signals that come from a common source section) but also how the brain arbitrates between integration and segregation in the face of uncertainty about the causal structure of the world. At small spatial disparities, the perceived location of an auditory event (e.g., the voice of a puppeteer) shifts toward the location of a temporally correlated but spatially displaced visual event (e.g., the facial movements of the puppet) and vice versa depending on the relative auditory and visual reliabilities as described in the forced fusion section.[25] This spatial biasing (i.e., the ventriloquist effect) breaks down or is at least attenuated at large spatial disparities and audiovisual asynchronies when it is unlikely that auditory and visual signals are caused by a common source.[70–74]

p0185 Initial modeling approaches introduced coupling priors to allow signals from different senses to bias each other without being integrated into one single unified percept.[75,76] More recently, Körding et al.[7] (and simultaneously Sato et al.[77]) proposed a Bayesian causal inference model that explicitly models the potential causal structures (i.e., common source or independent sources) that could have generated the sensory signals. Fig. 5.1C shows the generative model for Bayesian causal inference in an audiovisual spatial ventriloquist paradigm and localization task.

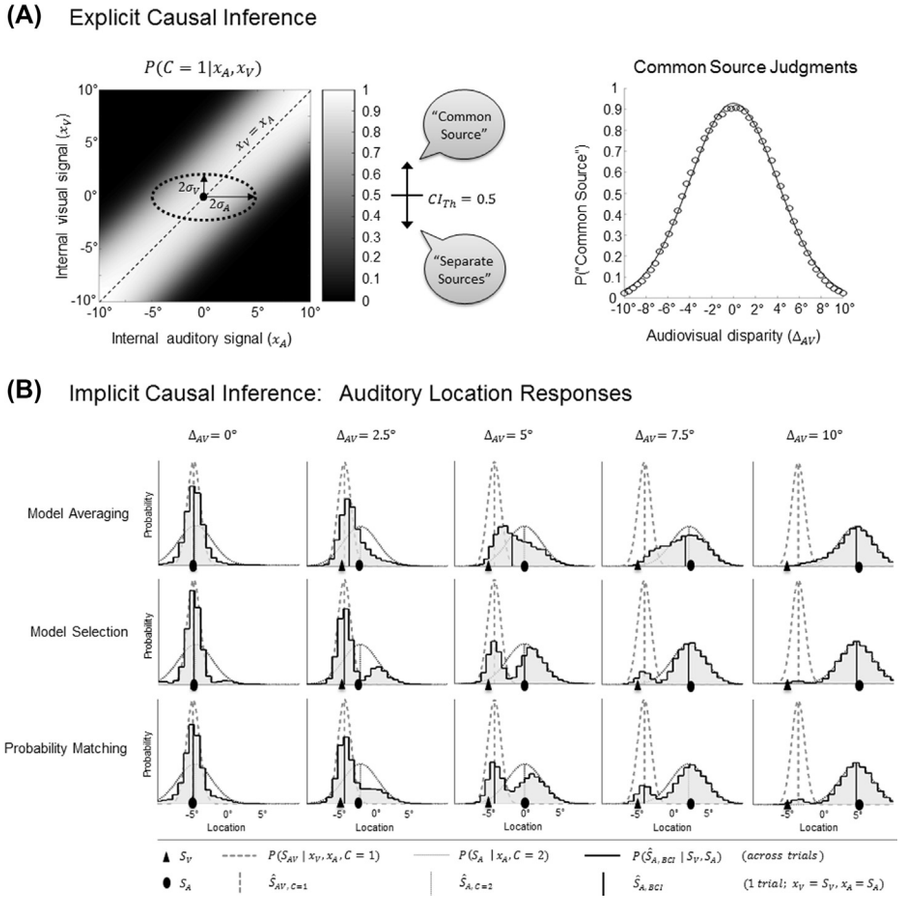
p0190 The generative model of Bayesian causal inference assumes that common ($C = 1$) or independent ($C = 2$) sources are determined by sampling from a binomial distribution with $P(C = 1)$ equal to the common-source prior P_{common}. The common source prior thus quantifies the observers' "unity assumption"[78] or prior tendency to integrate signals from different sensory modalities into one unified percept.

p0195 For a common source, the "true" location S_{AV} is drawn from the spatial prior distribution $N(\mu_{prior}, \sigma^2_{prior})$. For two independent causes, the "true" auditory (S_A) and visual (S_V) locations are drawn independently from this spatial prior distribution. The spatial prior distribution models an observer's prior expectations of where events may happen (see Combining information from a single sensory channel with prior knowledge section). For instance, we can model a central bias or expectation that events happen in the center of the visual field[14,15] by setting $\mu_{prior} = 0°$ and adjusting its strength in terms of the variance σ^2_{prior}.

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.



f0020 **FIGURE 5.3 Explicit and Implicit Bayesian Causal Inference.** (A) Explicit causal inference. The posterior probability of a common source $P(C = 1|x_A,x_V)$ is shown as a function of the internal auditory and visual signals (x_A and x_V). It decreases for increasing spatial disparities between the internal audiovisual signals. The observer is assumed to report a common source if the posterior probability for a common source is greater than a threshold CI_{th} (e.g., if $P(C = 1|x_A,x_V) > 0.5$). Critically, even if the true auditory and visual source locations are identical (i.e., $S_A = S_V$), the internal visual and auditory signals can differ because of internal and external noise (e.g., the area circumscribed by the *dashed black circle* covers 95% of the bivariate Gaussian probability distribution $P(x_A,x_V|S_A = 0°,S_V = 0°)$). Right panel: probability of a common source judgment (across trials) as a function of spatial disparity Δ_{AV} between the auditory and visual sources (S_A and S_V) as predicted by the Bayesian causal inference model (see text). (B) Implicit causal inference. Auditory location responses: simulated auditory location responses as a function of audiovisual spatial disparity (Δ_{AV}, columns 1 to 5) according to Bayesian causal inference for the three decision functions: model averaging (top row), model selection (middle row), and probability matching (bottom row). The black triangles indicate the true visual source location S_V and the black disks the true auditory source location S_A. For one trial per panel with $x_A = S_A$ and $x_V = S_V$: the dashed lines show the audiovisual posterior probability distributions $P(S_{AV}|x_A,x_V,C = 1)$ and audiovisual spatial estimates $\widehat{S}_{AV,C=1}$ (i.e., maximum a posteriori estimates; vertical lines) for the forced fusion model component. The dotted lines show the auditory posterior probability distributions $P(S_A|x_A,C = 2)$ and auditory spatial estimates $\widehat{S}_{A,C=2}$ for the full segregation model component. Finally, the *vertical solid lines* indicate the Bayesian causal inference estimate $\widehat{S}_{A,\,BCI}$. The solid lines

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

p0200 Finally, exactly as in the unisensory and the forced fusion cases, noise is introduced independently for each sensory modality by drawing the sensory inputs x_A and x_V independently from normal distributions centered on the true auditory (or visual) locations with parameter σ_A (or σ_V). Thus, σ_A and σ_V define the noise (i.e., reliability) of the inputs in each sensory modality.

p0205 In total, the generative model includes the following free parameters: the common-source prior P_{common}, the spatial prior standard deviation σ_{prior}, the auditory standard deviation σ_A, and the visual standard deviation σ_V.

p0210 Given this probabilistic generative model, the observer needs to infer the causal structure that has generated the sensory inputs (i.e., common source or causal judgment) and the location of the auditory and/or visual inputs (i.e., spatial localization task). Critically, as we will see below, an observer's spatial estimates inherently depend on his strategy of how to deal with his uncertainty about the underlying causal structure. In other words, the observer's implicit causal inference codetermines his spatial estimate during a localization task.

p0215 The posterior probability of the underlying causal structure can be inferred by combining the common-source prior with the sensory evidence according to Bayes' rule:[7]

$$P(C = 1|x_A, x_V) = \frac{P(x_A, x_V|C = 1) * P_{common}}{P(x_A, x_V)} \tag{5.10}$$

p0220 In explicit causal inference tasks (e.g., common source or congruency judgments), observers may thus report common or independent sources by applying a fixed threshold (e.g., $CI_{Th} = 0.5$) to the posterior probability of a common source:

$$\widehat{C} = \begin{cases} 1 & if \quad P(C = 1|x_A, x_V) \geq CI_{Th} \\ 2 & if \quad P(C = 1|x_A, x_V) < CI_{Th} \end{cases} \tag{5.11}$$

p0225 As expected and shown in Fig. 5.3A, the posterior probability for a common source decreases with increasing spatial disparity between the auditory and visual signals. Indeed, numerous studies have demonstrated that participants are less likely to perceive signals as coming from a common source for large intersensory conflicts such as audiovisual spatial disparity or temporal asynchrony.[62–64,70–73,79,80]

p0230 Critically, the estimate of the auditory and visual source location needs to be formed depending on the underlying causal structure: in the case of a known common source ($C = 1$), the optimal estimate of the audiovisual location is a reliability-weighted average of the auditory and visual percepts and the spatial prior (i.e., this is the forced fusion estimate

delineating the gray shaded area define the probability distributions (i.e., normalized histograms) of the Bayesian causal inference estimates across many trials $P\left(\widehat{S}_{A,BCI}|S_A, S_V\right)$. The distributions were generated from 10,000 randomly sampled x_A, x_V for each combination of S_A, S_V, with the parameters for visual noise: $\sigma_V = 1°$, auditory noise: $\sigma_A = 2.5°$, central spatial prior distribution: $\mu_{prior} = 0°$ and $\sigma_{prior} = 10°$, and common source prior: $P_{common} = 0.5$ (n.b. the same parameter values were used in panel A). *Adapted from Wozny DR, Beierholm UR, Shams L. Probability matching as a computational strategy used in perception.* PLoS Comput Biol. *2010;6(8).*

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

of Forced fusion: integrating sensory signals that come from a common source section, Eq. (5.6), with addition of the spatial prior):

$$\widehat{S}_{AV,C=1} = \frac{\frac{x_A}{\sigma_A^2} + \frac{x_V}{\sigma_V^2} + \frac{\mu_{prior}}{\sigma_{prior}^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2} + \frac{1}{\sigma_{prior}^2}} \tag{5.12}$$

p0235 In the case of known independent sources ($C = 2$), the optimal estimates of the auditory and visual signal locations (for the auditory and visual location report, respectively) are independent from each other (i.e., the so-called full segregation estimates).

$$\widehat{S}_{A,C=2} = \frac{\frac{x_A}{\sigma_A^2} + \frac{\mu_{prior}}{\sigma_{prior}^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_{prior}^2}} \text{ and } \widehat{S}_{V,C=2} = \frac{\frac{x_V}{\sigma_V^2} + \frac{\mu_{prior}}{\sigma_{prior}^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_{prior}^2}} \tag{5.13}$$

p0240 Critically, the observer does not know the underlying causal structure and hence needs to provide a final estimate of the auditory and visual locations that account for this causal uncertainty. More specifically, the observer can combine the estimates under the two causal structures using various decision functions such as "model averaging," "model selection," or "probability matching,"[81] as described below.

p0245 According to the "model averaging" strategy, the observer accounts for his causal uncertainty by combining the integrated, forced fusion spatial estimate with the segregated, task-relevant unisensory spatial estimate (i.e., either auditory or visual; whichever needs to be reported) weighted in proportion to the posterior probability of the underlying causal structures. This strategy minimizes the error about the spatial estimates under the assumption of a squared loss function.[7]

$$\widehat{S}_A = P(C = 1|x_A, x_V) * \widehat{S}_{AV,\,C=1} + (1 - P(C = 1|x_A, x_V)) * \widehat{S}_{A,\,C=2} \tag{5.14}$$

$$\widehat{S}_V = P(C = 1|x_A, x_V) * \widehat{S}_{AV,\,C=1} + (1 - P(C = 1|x_A, x_V)) * \widehat{S}_{V,\,C=2} \tag{5.15}$$

p0250 According to the "model selection" strategy, the observer reports the auditory ($\widehat{S}_A$) or visual ($\widehat{S}_V$) spatial estimate selectively from the more likely causal structure. This strategy minimizes the error about the inferred causal structures, as well as the error about the spatial estimates given the inferred causal structures.

$$\widehat{S}_A = \begin{cases} \widehat{S}_{AV,C=1} & \text{if } P(C = 1|x_A, x_V) \geq 0.5 \\ \widehat{S}_{A,C=2} & \text{if } P(C = 1|x_A, x_V) < 0.5 \end{cases} \tag{5.16}$$

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

$$\widehat{S}_V = \begin{cases} \widehat{S}_{AV,C=1} & if \quad P(C=1|x_A,x_V) \geq 0.5 \\ \widehat{S}_{V,C=2} & if \quad P(C=1|x_A,x_V) < 0.5 \end{cases} \tag{5.17}$$

p0255 According to "probability matching," the observer reports the spatial estimate of one causal structure stochastically selected in proportion to its posterior probability.

$$\widehat{S}_A = \begin{cases} \widehat{S}_{AV,C=1} & if \quad P(C=1|x_A,x_V) \geq \alpha \\ \widehat{S}_{A,C=2} & if \quad P(C=1|x_A,x_V) < \alpha \end{cases} \text{ with } \alpha \sim Uniform(0,1) \tag{5.18}$$

$$\widehat{S}_V = \begin{cases} \widehat{S}_{AV,C=1} & if \quad P(C=1|x_A,x_V) \geq \alpha \\ \widehat{S}_{V,C=2} & if \quad P(C=1|x_A,x_V) < \alpha \end{cases} \text{ with } \alpha \sim Uniform(0,1) \tag{5.19}$$

p0260 As illustrated in Fig. 5.3B, Bayesian causal inference transitions gracefully between sensory integration and segregation as a function of intersensory conflict irrespective of the specific decision function. In other words, while the forced fusion model allows only for a linear combination of the sensory signals ($\widehat{S}_{AV,C=1}$ in Fig. 5.3B), Bayesian causal inference models ($\widehat{S}_{A,\,BCI}$) combine sensory signals nonlinearly as a function of intersensory conflict. They predominantly integrate sensory signals approximately in line with forced fusion models, when the conflict is small, but attenuate integration for large conflicts. Numerous studies since the inception of multisensory integration as a research field in its own right have provided qualitative evidence for the computational principles governing Bayesian causal inference. For instance, several studies have demonstrated an inverted U-shape function for % perceived synchronous or the McGurk effect as a function of audiovisual synchrony of speech signals.[60,62,63,67]

p0265 Over the past decade, accumulating research has also quantitatively compared human behavior with the predictions of Bayesian causal inference in a range of tasks including audiovisual spatial localization,[7,15,74,77,79–86] audiovisual temporal discrimination,[86–88] visual-vestibular heading estimation,[89] audiovisual speech recognition,[90] audiovisual distance perception,[91] and audiovisuo–tactile numerosity judgments.[92] In the following, we discuss the role of (1) reliability of the sensory inputs, (2) the common source prior, and (3) the decision function in Bayesian causal inference.

p0270 To investigate the influence of sensory reliability on how human observers arbitrate between sensory integration and segregation, Rohe and Noppeney[79] presented participants with auditory and visual spatial signals at multiple spatial disparities and visual reliabilities. In a dual task, observers performed Bayesian causal inference implicitly for auditory spatial localization and explicitly for common source judgment. The study showed that visual reliability shapes multisensory integration not only by determining the relative sensory weights but also by defining the spatial integration window. As expected by Bayesian causal inference, highly reliable visual signals sensitized observers to audiovisual disparity thereby sharpening the spatial integration window.[79]

p0275 In addition to bottom-up sensory signals, Bayesian causal inference depends on the so-called "common source prior," embodying an observer's prior expectations that two signals are caused by a common source. This raises the question whether these common source

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

priors are hardwired in an individual, specifically for a particular task and stimulus characteristics. For instance, in a conversational setting with a single speaker, we should be more inclined to integrate his/her facial movements with the syllables he/she is uttering for improved speech comprehension. By contrast, in a busy pub where we are bombarded with many conflicting auditory and visual speech signals, unconstrained information integration would be detrimental. In a first study, Odegaard and Shams[86] showed that common source priors are relatively stable across time (also see Beierholm et al.[85]), yet task-specific. More specifically, they did not generalize from a spatial ventriloquism task to a double flash illusion task. Yet, in a follow-up study where they dynamically manipulated the probability of audiovisual signals being synchronous and colocated, in a ventriloquist paradigm, they demonstrated that observers dynamically adapt their common source priors to the environmental statistics.[74] Indeed, dynamic adjustment of common source priors had also previously been shown during audiovisual speech perception.[93–95]

p0280 Finally, Wozny et al.[81] investigated in a large cohort of more than 100 observers, whether observers are more likely to use model averaging, model selection, or probability matching as decisional functions in Bayesian causal inference. Surprisingly, they demonstrated that human observers predominantly use probability matching in audiovisual spatial localization. While probability matching may be thought of as being suboptimal for static environments, humans have been shown to use this strategy in a variety of cognitive tasks (e.g., reward learning[96,97]). The authors proposed that probability matching may be a useful strategy to explore potential causal structures in a dynamic environment. In summary, accumulating psychophysical research has shown that human perception is governed qualitatively and to some extent quantitatively by the principles of Bayesian causal inference, raising the question of how the brain may compute Bayesian causal inference.

p0285 At the neural level, extensive neurophysiological and neuroimaging evidence has demonstrated that multisensory integration, as indexed by multisensory response enhancement or suppression relative to the unisensory responses, depends on a temporal and spatial window of integration.[98,99] Spatial windows of integration may be related to neuronal receptive field properties. By contrast, temporal windows of integration may rely on computation of temporal correlations (e.g., see recent model using the Hassenstein-Reichardt detector[100]) and have recently been associated with brain oscillations.[101–103] Models for the neural implementations of Bayesian causal inference have been proposed, but their biological plausibility still needs to be shown.[104–107]

p0290 At the neural systems level, two recent neuroimaging studies by Rohe and Noppeney[82,83] investigated how the brain accomplishes Bayesian causal inference by combining psychophysics, fMRI, Bayesian modeling, and multivariate decoding. On each trial, participants localized audiovisual signals that varied in spatial discrepancy and visual reliability. The studies demonstrated that the brain computes Bayesian causal inference by encoding multiple spatial estimates across the cortical hierarchy. At the bottom of the hierarchy, in auditory and visual cortical areas, location is represented on the basis that the two signals are generated by independent sources (= segregation). At the next stage, in the posterior intraparietal sulcus, location is estimated under the assumption that the two signals are from a common source (= forced fusion). It is only at the top of the hierarchy, in the anterior intraparietal sulcus, that the uncertainty about whether signals are generated by common or independent sources is taken into account. As predicted by Bayesian causal inference, the final location

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

is computed by combining the segregation and the forced fusion estimates, weighted by the posterior probabilities of common and independent sources.

s0030

Conclusions

p0295 Bayesian models of perceptual inference define how an observer should integrate uncertain sensory signals to provide an accurate and reliable percept of our environment. They thus set a benchmark of an ideal observer against which human perceptual performance can be compared. Forced fusion models and psychophysical studies have highlighted that human observers integrate sensory signals that come from a common source weighted approximately in proportion to their relative reliabilities. More recent models of Bayesian causal inference account for an observer's uncertainty about the world's causal structure by explicitly modeling whether sensory signals come from common or independent sources. A final Bayesian causal inference estimate is then obtained by combining the estimates under the assumptions of common or independent sources according to various decision functions. Accumulating psychophysical and neuroimaging evidence has recently suggested that human observers perform spatial localization and speech recognition tasks in line with the principles of Bayesian causal inference.

Acknowledgments

This research was funded by ERC-2012-StG_20111109 multsens.

References

1. Honkanen A, Immonen EV, Salmela I, Heimonen K, Weckstrom M. Insect photoreceptor adaptations to night vision. *Philos Trans R Soc Lond B Biol Sci*. 2017;372(1717).
2. Shams L, Beierholm UR. Causal inference in perception. *Trends Cognit Sci*. 2010;14(9):425–432.
3. Barlow HB. Retinal noise and absolute threshold. *J Opt Soc Am*. 1956;46(8):634–639.
4. Stevens CF. Neurotransmitter release at central synapses. *Neuron*. 2003;40(2):381–388.
5. Gregory RL. Knowledge in perception and illusion. *Philos Trans R Soc Lond B Biol Sci*. 1997;352(1358):1121–1127.
6. Yuille AL, Bulthoff HH. Bayesian decision theory and psychophysics. In: David CK, Whitman R, eds. *Perception as Bayesian Inference*. Cambridge University Press; 1996:123–161.
7. Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. *PLoS One*. 2007;2(9):e943.
8. Geisler WS, Diehl RL. Bayesian natural selection and the evolution of perceptual systems. *Philos Trans R Soc Lond B Biol Sci*. 2002;357(1420):419–448.
9. Gopnik A, Tenenbaum JB. Bayesian networks, Bayesian learning and cognitive development. *Dev Sci*. 2007;10(3):281–287.
10. Roach NW, McGraw PV, Whitaker DJ, Heron J. Generalization of prior information for rapid Bayesian time estimation. *Proc Natl Acad Sci U S A*. 2017;114(2):412–417.
11. Di Luca M, Rhodes D. Optimal perceived timing: integrating sensory information with dynamically updated expectations. *Sci Rep*. 2016;6:28563.
12. Mamassian P, Landy MS. Interaction of visual prior constraints. *Vis Res*. 2001;41(20):2653–2668.
13. Jacobs RA. Optimal integration of texture and motion cues to depth. *Vis Res*. 1999;39(21):3621–3629.
14. Kerzel D. Memory for the position of stationary objects: disentangling foveal bias and memory averaging. *Vis Res*. 2002;42(2):159–167.

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

15. Odegaard B, Wozny DR, Shams L. Biases in visual, auditory, and audiovisual perception of space. *PLoS Comput Biol*. 2015;11(12):e1004649.
16. Weiss Y, Simoncelli EP, Adelson EH. Motion illusions as optimal percepts. *Nat Neurosci*. 2002;5(6):598–604.
17. Stocker AA, Simoncelli EP. Noise characteristics and prior expectations in human visual speed perception. *Nat Neurosci*. 2006;9(4):578–585.
18. Girshick AR, Landy MS, Simoncelli EP. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat Neurosci*. 2011;14(7):926–932.
19. Jazayeri M, Shadlen MN. Temporal context calibrates interval timing. *Nat Neurosci*. 2010;13(8):1020–1026.
20. Kording KP, Wolpert DM. Bayesian integration in sensorimotor learning. *Nature*. 2004;427(6971):244–247.
21. van Bergen RS, Ma WJ, Pratte MS, Jehee JF. Sensory uncertainty decoded from visual cortex predicts behavior. *Nat Neurosci*. 2015;18(12):1728–1730.
22. Oruc I, Maloney LT, Landy MS. Weighted linear cue combination with possibly correlated error. *Vis Res*. 2003;43(23):2451–2468.
23. Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*. 2002;415(6870):429–433.
24. Rohde M, van Dam LCJ, Ernst MO. Statistically optimal multisensory cue integration: a practical tutorial. *Multisensory Res*. 2016;29(4–5):279–317.
25. Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol*. 2004;14(3):257–262.
26. Kingdom FAA, Prins N. Chapter 4 - Psychometric Functions*. In: *Psychophysics*. 2nd ed. San Diego: Academic Press; 2016:55–117.
27. Acuna DE, Berniker M, Fernandes HL, Kording KP. Using psychophysics to ask if the brain samples or maximizes. *J Vis*. 2015;15(3).
28. Young MJ, Landy MS, Maloney LT. A perturbation analysis of depth perception from combinations of texture and motion cues. *Vis Res*. 1993;33(18):2685–2696.
29. Gepshtein S, Burge J, Ernst MO, Banks MS. The combination of vision and touch depends on spatial proximity. *J Vis*. 2005;5(11):1013–1023.
30. Raposo D, Sheppard JP, Schrater PR, Churchland AK. Multisensory decision-making in rats and humans. *J Neurosci*. 2012;32(11):3726–3735.
31. Sheppard JP, Raposo D, Churchland AK. Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans. *J Vis*. 2013;13(6).
32. Helbig HB, Ernst MO. Optimal integration of shape information from vision and touch. *Exp Brain Res*. 2007;179(4):595–606.
33. Hartcher-O'Brien J, Di Luca M, Ernst MO. The duration of uncertain times: audiovisual information about intervals is integrated in a statistically optimal fashion. *PLoS One*. 2014;9(3):e89339.
34. Mendonca C, Santos JA, Lopez-Moliner J. The benefit of multisensory integration with biological motion signals. *Exp Brain Res*. 2011;213(2–3):185–192.
35. Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE. Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci*. 2011;15(1):146–154.
36. Gu Y, Angelaki DE, Deangelis GC. Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci*. 2008;11(10):1201–1210.
37. Morgan ML, Deangelis GC, Angelaki DE. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*. 2008;59(4):662–673.
38. Ohshiro T, Angelaki DE, DeAngelis GC. A normalization model of multisensory integration. *Nat Neurosci*. 2011;14(6):775–782.
39. Ohshiro T, Angelaki DE, DeAngelis GC. A neural signature of divisive normalization at the level of multisensory integration in primate cortex. *Neuron*. 2017;95(2):399–411 e398.
40. Beauchamp MS, Pasalar S, Ro T. Neural substrates of reliability-weighted visual-tactile multisensory integration. *Front Syst Neurosci*. 2010;4:25.
41. Nath AR, Beauchamp MS. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *J Neurosci*. 2011;31(5):1704–1714.
42. Helbig HB, Ernst MO, Ricciardi E, et al. The neural mechanisms of reliability weighted integration of shape information from vision and touch. *Neuroimage*. 2012;60(2):1063–1072.

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

43. Battaglia PW, Jacobs RA, Aslin RN. Bayesian integration of visual and auditory signals for spatial localization. *J Opt Soc Am A*. 2003;20(7):1391−1397.
44. Burr D, Banks MS, Morrone MC. Auditory dominance over vision in the perception of interval duration. *Exp Brain Res*. 2009;198(1):49−57.
45. Maiworm M, Röder B. Suboptimal auditory dominance in audiovisual integration of temporalcues. *Tsinghua Sci Technol*. 2011;16(2):121−132.
46. Fetsch CR, Turner AH, DeAngelis GC, Angelaki DE. Dynamic reweighting of visual and vestibular cues during self-motion perception. *J Neurosci*. 2009;29(49):15601−15612.
47. Butler JS, Smith ST, Campos JL, Bulthoff HH. Bayesian integration of visual and vestibular signals for heading. *J Vis*. 2010;10(11):23.
48. Prsa M, Gale S, Blanke O. Self-motion leads to mandatory cue fusion across sensory modalities. *J Neurophysiol*. 2012;108(8):2282−2291.
49. Rosas P, Wagemans J, Ernst MO, Wichmann FA. Texture and haptic cues in slant discrimination: reliability-based cue weighting without statistically optimal cue combination. *J Opt Soc Am A*. 2005;22(5):801−809.
50. Bentvelzen A, Leung J, Alais D. Discriminating audiovisual speed: optimal integration of speed defaults to probability summation when component reliabilities diverge. *Perception*. 2009;38(7):966−987.
51. Gepshtein S, Banks MS. Viewing geometry determines how vision and haptics combine in size perception. *Curr Biol*. 2003;13(6):483−488.
52. Battaglia PW, Kersten D, Schrater PR. How haptic size sensations improve distance perception. *PLoS Comput Biol*. 2011;7(6):e1002080.
53. Jacobs RA. What determines visual cue reliability? *Trends Cognit Sci*. 2002;6(8):345−350.
54. Gori M, Sciutti A, Burr D, Sandini G. Direct and indirect haptic calibration of visual size judgments. *PLoS One*. 2011;6(10):e25599.
55. Wozny DR, Shams L. Recalibration of auditory space following milliseconds of cross-modal discrepancy. *J Neurosci*. 2011;31(12):4607−4612.
56. Ernst MO. Optimal multisensory integration: assumptions and limits. In: Stein BE, ed. *The New Handbook of Multisensory Processes*. Cambridge, Massachusetts: MIT Press; 2012:527−544.
57. Warren DH, Welch RB, McCarthy TJ. The role of visual-auditory "compellingness" in the ventriloquism effect: implications for transitivity among the spatial senses. *Percept Psychophys*. 1981;30(6):557−564.
58. Bishop CW, Miller LM. Speech cues contribute to audiovisual spatial integration. *PLoS One*. 2011;6(8):e24016.
59. Kanaya S, Yokosawa K. Perceptual congruency of audio-visual speech affects ventriloquism with bilateral visual stimuli. *Psychon Bull Rev*. 2011;18(1):123−128.
60. Lee H, Noppeney U. Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proc Natl Acad Sci U S A*. 2011;108(51). E1441-1450.
61. Lee H, Noppeney U. Temporal prediction errors in visual and auditory cortices. *Curr Biol*. 2014;24(8):R309−R310.
62. van Wassenhove V, Grant KW, Poeppel D. Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*. 2007;45(3):598−607.
63. Soto-Faraco S, Alsius A. Deconstructing the McGurk-MacDonald illusion. *J Exp Psychol Hum Percept Perform*. 2009;35(2):580−587.
64. Stevenson RA, Fister JK, Barnett ZP, Nidiffer AR, Wallace MT. Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Exp Brain Res*. 2012;219(1):121−137.
65. Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ. The effect of prior visual information on recognition of speech and sounds. *Cerebr Cortex*. 2008;18(3):598−609.
66. Adam R, Noppeney U. Prior auditory information shapes visual category-selectivity in ventral occipito-temporal cortex. *Neuroimage*. 2010;52(4):1592−1602.
67. Maier JX, Di Luca M, Noppeney U. Audiovisual asynchrony detection in human speech. *J Exp Psychol Hum Percept Perform*. 2011;37(1):245−256.
68. Parise CV, Spence C. 'When birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One*. 2009;4(5):e5664.
69. Parise CV, Spence C, Ernst MO. When correlation implies causation in multisensory integration. *Curr Biol*. 2012;22(1):46−49.

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

70. Slutsky DA, Recanzone GH. Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*. 2001;12(1):7–10.
71. Lewald J, Guski R. Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cogn Brain Res*. 2003;16(3):468–478.
72. Hairston WD, Wallace MT, Vaughan JW, Stein BE, Norris JL, Schirillo JA. Visual localization ability influences cross-modal bias. *J Cogn Neurosci*. 2003;15(1):20–29.
73. Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA. Unifying multisensory signals across time and space. *Exp Brain Res*. 2004;158(2):252–258.
74. Odegaard B, Wozny DR, Shams L. A simple and efficient method to enhance audiovisual binding tendencies. *PeerJ*. 2017;5:e3143.
75. Bresciani JP, Dammeier F, Ernst MO. Vision and touch are automatically integrated for the perception of sequences of events. *J Vis*. 2006;6(5):554–564.
76. Roach NW, Heron J, McGraw PV. Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proc Biol Sci*. 2006;273(1598):2159–2168.
77. Sato Y, Toyoizumi T, Aihara K. Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Comput*. 2007;19(12):3335–3355.
78. Chen YC, Spence C. Assessing the role of the 'unity assumption' on multisensory integration: a review. *Front Psychol*. 2017;8:445.
79. Rohe T, Noppeney U. Sensory reliability shapes perceptual inference via two mechanisms. *J Vis*. 2015;15(5):22.
80. Bosen AK, Fleming JT, Brown SE, Allen PD, O'Neill WE, Paige GD. Comparison of congruence judgment and auditory localization tasks for assessing the spatial limits of visual capture. *Biol Cybern*. 2016;110(6):455–471.
81. Wozny DR, Beierholm UR, Shams L. Probability matching as a computational strategy used in perception. *PLoS Comput Biol*. 2010;6(8).
82. Rohe T, Noppeney U. Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biol*. 2015;13(2):e1002073.
83. Rohe T, Noppeney U. Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Curr Biol*. 2016;26(4):509–514.
84. Natarajan R, Murray I, Shams L, Zemel RS. Characterizing response behavior in multisensory perception with conflicting cues. In: Koller D, Schuurmans D, Bengio Y, Bottou L, eds. *Advances in Neural Information Processing Systems*. Vol. 21. MIT Press; 2009:1153–1160.
85. Beierholm UR, Quartz SR, Shams L. Bayesian priors are encoded independently from likelihoods in human multisensory perception. *J Vis*. 2009;9(5), 23 21-29.
86. Odegaard B, Shams L. The brain's tendency to bind audiovisual signals is stable but not general. *Psychol Sci*. 2016;27(4):583–591.
87. Magnotti JF, Ma WJ, Beauchamp MS. Causal inference of asynchronous audiovisual speech. *Front Psychol*. 2013;4:798.
88. McGovern DP, Roudaia E, Newell FN, Roach NW. Perceptual learning shapes multisensory causal inference via two distinct mechanisms. *Sci Rep*. 2016;6:24673.
89. de Winkel KN, Katliar M, Bulthoff HH. Causal inference in multisensory heading estimation. *PLoS One*. 2017;12(1):e0169676.
90. Magnotti JF, Beauchamp MS. A causal inference model explains perception of the McGurk effect and other incongruent audiovisual speech. *PLoS Comput Biol*. 2017;13(2):e1005229.
91. Mendonca C, Mandelli P, Pulkki V. Modeling the perception of audiovisual distance: Bayesian causal inference and other models. *PLoS One*. 2016;11(12):e0165391.
92. Wozny DR, Beierholm UR, Shams L. Human trimodal perception follows optimal statistical inference. *J Vis*. 2008;8(3), 24 21-11.
93. Nahorna O, Berthommier F, Schwartz JL. Binding and unbinding the auditory and visual streams in the McGurk effect. *J Acoust Soc Am*. 2012;132(2):1061–1077.
94. Nahorna O, Berthommier F, Schwartz JL. Audio-visual speech scene analysis: characterization of the dynamics of unbinding and rebinding the McGurk effect. *J Acoust Soc Am*. 2015;137(1):362–377.
95. Gau R, Noppeney U. How prior expectations shape multisensory perception. *Neuroimage*. 2016;124(Pt A):876–886.

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter TNQ Books and Journals Pvt Ltd. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

96. Erev I, Roth AE. Maximization, learning, and economic behavior. *Proc Natl Acad Sci U S A*. 2014;111(Suppl. 3):10818–10825.
97. Vul E, Goodman N, Griffiths TL, Tenenbaum JB. One and done? Optimal decisions from very few samples. *Cogn Sci*. 2014;38(4):599–637.
98. Meredith MA, Nemitz JW, Stein BE. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci*. 1987;7(10):3215–3229.
99. Meredith MA, Stein BE. Spatial determinants of multisensory integration in cat superior colliculus neurons. *J Neurophysiol*. 1996;75(5):1843–1857.
100. Parise CV, Ernst MO. Correlation detection as a general mechanism for multisensory integration. *Nat Commun*. 2016;7:11543.
101. Cecere R, Rees G, Romei V. Individual differences in alpha frequency drive crossmodal illusory perception. *Curr Biol*. 2015;25(2):231–235.
102. Samaha J, Postle BR. The speed of alpha-band oscillations predicts the temporal resolution of visual perception. *Curr Biol*. 2015;25(22):2985–2990.
103. Thakur B, Mukherjee A, Sen A, Banerjee A. A dynamical framework to relate perceptual variability with multisensory information processing. *Sci Rep*. 2016;6:31280.
104. Ma WJ, Rahmati M. Towards a neural implementation of causal inference in cue combination. *Multisensory Res*. 2013;26(1–2):159–176.
105. Spratling MW. A neural implementation of Bayesian inference based on predictive coding. *Connect Sci*. 2016;28(4):346–383.
106. Yu Z, Chen F, Dong J, Dai Q. Sampling-based causal inference in cue combination and its neural implementation. *Neurocomputing*. 2016;175:155–165.
107. Cuppini C, Shams L, Magosso E, Ursino M. A biologically inspired neurocomputational model for audiovisual integration and causal inference. *Eur J Neurosci*. 2017;46(9):2481–2498.

10005-SATHIAN-9780128124925

These proofs may contain color figures. Those figures may print black and white in the final printed book if a color print product has not been planned. The color figures will appear in color in all electronic versions of this book.

SATHIAN: 05

Non-Print Items

Abstract
Normative Bayesian models of perceptual inference define how observers should combine uncertain information across multiple sensory channels and prior knowledge to obtain the most reliable percept of our environment. In this review, we first introduce forced fusion models that describe how observers integrate sensory signals along with prior knowledge approximately weighted in proportion to their relative reliabilities. Yet, these models describe only the special case of mandatory integration that applies when signals come necessarily from a common source; they cannot model situations where signals can come from common or independent sources. In these more naturalistic situations, observers should integrate signals from common sources but segregate those from independent sources. Recent hierarchical models of Bayesian causal inference solve this so-called causal inference problem by explicitly modeling the world's causal structure (i.e., common or independent sources). To account for observers' uncertainty about the world's causal structure, a final Bayesian causal inference estimate is then obtained by combining the estimates under the assumptions of common or independent sources according to various decision functions (e.g., model averaging). Growing psychophysical and neuroimaging evidence suggests that human observers arbitrate between sensory integration and segregation in line with the principles of Bayesian causal inference.

Keywords:
Bayesian causal inference; Maximum likelihood estimation; Multisensory perception; Prior probability distribution; Probabilistic computational models; Reliability-weighted integration.